AMENDMENTS TO THE CLAIMS

The listing of claims below replaces all prior versions of claims in the application.

1. (Currently Amended): A robotics visual and auditory system comprising[[;]]:

an auditory module for collecting external sounds by at least a pair of microphones, and determining a direction of at least one speaker by sound source separation and localization by grouping based on pitch extraction and harmonic sounds from the sound signals collected by the microphones;

a face module for taking images of a robot's front with a camera, and identifying each speaker;

a speech recognition part for conducting speech recognition;

wherein the speech recognition part includes:

a plurality of acoustic models[[,]];

a speech recognition engine for ~~executing speech recognition processes to~~ processing a plurality of separated sound signals from respective sound sources to execute speech recognition processes by using the acoustic models, and

a selector for integrating a plurality of speech recognition process results obtained by the speech recognition engine ~~speech recognition process,~~ and selecting any one of speech recognition process results[[,]];

wherein[[,]] in order to respond to cases ~~the case~~ where a plurality of speakers speak to [[said]] the robot from different directions with respect to the robot's front direction as [[the]] a

base, the acoustic models are provided with respect to each speaker and each direction ~~so to respond each direction,~~;

wherein the speech recognition engine uses each of [[said]] the acoustic models ~~separately~~ for one sound signal separated by sound source separation[[,]] and executes [[said]] the speech recognition ~~process~~ processes in parallel.

2. (Currently Amended): A robotics visual and auditory system as set forth in Claim 1, wherein the selector calculates [[the]] a cost function value[[,]] based on the recognition result by the speech recognition process and the speaker's direction upon integrating the speech recognition process result~~, based on the recognition result by the speech recognition process and the speaker's direction,~~ and judges [[the]] a speech recognition process result having the maximum value of [[the]] a cost function as [[the]] a most reliable speech recognition result.

3. (Currently Amended): A robotics visual and auditory system as set forth in Claim 1 or Claim 2, wherein [[it]] the system is provided with a dialogue part to output the speech recognition process results selected by the selector to outside.

4. (Currently Amended): A robotics visual and auditory system comprising[[;]]:

an auditory module ~~which is provided at least with a pair of microphones to collect~~ for collecting external sounds by at least a pair of microphones, ~~and, based on sound signals from the microphones,~~ determining a direction of at least one speaker by sound source separation and

localization by grouping based on pitch extraction and harmonic sounds from the sound signals collected by the microphones, and extracting an auditory event[[,]];

a face module ~~which is provided a camera to take~~ for taking images of a robot's front with a camera, identifying ~~identify~~ each speaker, and ~~extracts his~~ extracting a face event from each speaker's face recognition and localization[[,]] based on images taken by the camera[[,]];

a motor control module ~~which is provided with a drive motor to rotate the~~ for rotating a robot in [[the]] a horizontal direction by a drive motor, and ~~extracts~~ extracting a motor event[[,]] based on a rotational position of the drive motor[[,]];

an association module ~~which determines~~ for determining each speaker's direction[[,]] based on directional information of sound source localization of the auditory event and face localization of the face event[[,]] from [[said]] the auditory, face, and motor events, ~~generates~~ generating an auditory stream and a face stream by respectively connecting ~~said events~~ the auditory event and the face event in the temporal direction using a Kalman filter for determinations, and further ~~generates~~ generating an association stream by associating ~~these streams, and~~ the auditory stream with the face stream,

an attention control module ~~which conduct~~ for conducting an attention control based on ~~said streams, and drive controls~~ the association stream, the auditory stream and the face stream and controlling the motor based on an action planning results accompanying the attention control, and

a speech recognition part for conducting speech recognitions:

wherein the speech recognition part includes:

a plurality of acoustic models;

a speech recognition engine for processing a plurality of separated sound signals from respective sound sources to execute speech recognition processes by using the acoustic models, and

a selector for integrating a plurality of speech recognition process results obtained by the speech recognition engine and selecting any one of speech recognition process results;

wherein in order ~~for the auditory module~~ to respond to cases ~~the case~~ where a plurality of speakers speak to [[said]] the robot from different directions with respect to a [[the]] robot's front direction as [[the]] a base, the acoustic models are provided with respect to [[in]] each speaker and each direction ~~so to respond each speaker, and each direction,~~;

wherein the auditory module collects sub-bands having interaural phase difference (IPD) or interaural intensity difference (IID) within a predetermined range by an active direction pass filter having a pass range which~~, according to auditory characteristics,~~ becomes minimum in [[the]] a frontal direction[[,]] and becomes larger as [[the]] an angle becomes wider to [[the]] left and right[[,]] on the basis of ~~based on~~ an accurate sound source directional information from the association module[[,]] and conducts sound source separation by restructuring a wave shape of a sound source[[,]];

wherein the speech recognition engine conducts speech ~~recognition~~ recognition by using a plurality of the acoustic models in parallel for one sound signal separated by sound source separation ~~using a plurality of the acoustic models~~, and

wherein the selector integrates speech recognition results from each acoustic model ~~by a~~

~~selector,~~ and judges [[the]] a most reliable speech recognition result among the speech recognition results.

5 (Currently Amended). A robotics visual and auditory system comprising[[;]]:

an auditory module ~~which is provided at least with a pair of microphones to collect~~ for collecting external sounds by at least a pair of microphones, ~~and, based on sound signals from the microphones, determines~~ determining a direction of at least one speaker by sound source separation and localization by grouping based on pitch extraction and harmonic sounds from the sound signals collected by the microphones, and extracting an auditory event[[,]];

a face module ~~which is provided a camera to take~~ for taking images of a robot's front by camera, ~~identifies~~ identifying each speaker, and extracting a ~~extracts his~~ face event from each speaker's face recognition and localization, based on images taken by the camera[[,]];

a stereo module ~~which extracts and localizes~~ for extracting and localizing a longitudinally long matter[[,]] based on a parallax extracted from images taken by a stereo camera[[,]] and ~~extracts~~ extracting a stereo event[[,]];

a motor control module ~~which is provided with a drive motor to rotate the~~ for rotating a robot in [[the]] a horizontal direction by a drive motor[[,]] and ~~extracts~~ extracting a motor event[[,]] based on a rotational position of the drive motor[[,]];

an association module ~~which determines~~ for determining each speaker's direction[[,]] based on directional information of sound source localization of the auditory event and face localization of the face event[[,]] from [[said]] the auditory, face, stereo, and motor events,

~~generates~~ generating an auditory stream, a face stream and a stereo visual stream by respectively connecting [[said]] auditory events, face events, and stereo events in [[the]] a temporal direction using a Kalman filter for determinations, and further generating ~~generates~~ an association stream by associating ~~these~~ the auditory stream with the face and stereo visual streams, [[and]]

an attention control module ~~which conducts~~ for conducting an attention control based on ~~said streams, and drive controls~~ the association stream, the auditory stream, the face stream and the stereo visual stream, and controlling the motor based on an action planning results accompanying the attention control, and

a speech recognition part for conducting speech recognitions:

wherein the speech recognition part includes:

a plurality of acoustic models;

a speech recognition engine for processing a plurality of separated sound signals from respective sound sources to execute speech recognition processes by using the acoustic models, and

a selector for integrating a plurality of speech recognition process results obtained by the speech recognition engine and selecting any one of speech recognition process results;

wherein in order ~~for the auditory module~~ to respond to cases ~~the case~~ where a plurality of speakers speak to [[said]] the robot from different directions with respect to a [[the]] robot's front direction as [[the]] a base, the acoustic models are provided with respect to [[in]] each speaker and each direction ~~so to respond each speaker, and each direction,~~;

wherein the auditory module collects sub-bands having interaural phase difference (IPD)

or interaural intensity difference (IID) within a predetermined range by an active direction pass filter having a pass range which~~, according to auditory characteristics,~~ becomes minimum in [[the]] a frontal direction[[,]] and becomes larger as [[the]] an angle becomes wider to [[the]] left and right[[,]] on the basis of ~~based on~~ an accurate sound source directional information from the association module[[,]] and conducts sound source separation by restructuring a wave shape of a sound source[[,]];

wherein the speech recognition engine conducts speech ~~recognition~~ recognition using a plurality of the acoustic models in parallel for one sound signal separated by sound source separation ~~using a plurality of the acoustic models~~, and

wherein the selector integrates speech recognition results from each acoustic model ~~by a selector,~~ and judges [[the]] a most reliable speech recognition result among the speech recognition results.

6. (Currently Amended): A robotics visual and auditory system as set forth in Claim 4 or Claim 5, ~~characterized in that;~~ wherein when the speech recognition by the auditory module failed, the attention control module is made up so as to collect speeches again from the microphones after [[with]] the microphones ~~and the camera turned~~ turn to the sound source direction of the sound signals, and to perform again speech recognition of the speech by the auditory module[[,]] based on the sound signals conducted sound source localization and sound source separation.

7. (Currently Amended): A robotics visual and auditory system as set forth in Claim 4 or Claim 5, ~~characterized in that;~~ wherein the auditory module refers to the face event from the face module upon performing the speech recognition.

8. (Currently Amended): A robotics visual and auditory system as set forth in Claim 5, ~~characterized in that;~~ wherein the auditory module refers to the stereo event from the stereo module upon performing the speech recognition.

9. (Currently Amended): A robotics visual and auditory system as set forth in Claim 5, ~~characterized in that;~~ wherein

the auditory module refers to the face event from the face module and the stereo event from the stereo module upon performing the speech recognition.

10. (Currently Amended): A robotics visual and auditory system as set forth in Claim 4 or Claim 5, wherein [[it]] the system is provided with a dialogue part to output the speech recognition results judged by the auditory module to outside.

11. (Original): A robotics visual and auditory system as set forth in Claim 4 or Claim 5, wherein a pass range of the active direction pass filter can be controlled for each frequency.

12. (Currently Amended): A robotics visual and auditory system as set forth in Claim 4

or Claim 5, wherein the selector calculates [[the]] a cost function value, upon integrating the speech recognition result, based on the recognition result by the speech recognition and the direction determined by the association module, and judges [[the]] a speech recognition process result having [[the]] a maximum value of the cost function as the most reliable speech recognition result.

13. (Currently Amended): A robotics visual and auditory system as set forth in Claim 4 or Claim 5, ~~characterized in that; it~~ wherein the selector recognizes [[the]] a speaker's name based on the acoustic model utilized to obtain speech recognition result.